EXAMINATION AND ANALYSIS OF RESIDUALS.
I. DIAGNOSTIC CHECKING OF RESIDUALS IN LINEAR REGRESSION WITHOUT INTERCEPT FOR DETECTING A SPECIAL TYPE OF HETEROSCEDASTICITY

BU-247-M

Abdossamad Hedayat

July, 1967

ABSTRACT

The main purpose of this paper is to use the Theil residuals for detecting a special type of heteroscedasticity in linear regression analysis by means of the peak test.

Biometrics Unit, Cornell University, Ithaca, New York.

EXAMINATION AND ANALYSIS OF RESIDUALS.
I. DIAGNOSTIC CHECKING OF RESIDUALS IN LINEAR REGRESSION WITHOUT INTERCEPT FOR DETECTING A SPECIAL TYPE OF HETEROSCEDASTICITY

BU-247-M

Abdossamad Hedayat

July, 1967

Introduction

Consider the simple linear model

(1) $$Y = X\beta + \epsilon$$

where

(a) Y represents an n-dimensional random vector,

(b) X is an n-dimensional column vector with known coefficients and consists of nonstochastic elements, or if not, are distributed independently of the error terms,

(c) β is an unknown scalar,

(d) ϵ is an n-dimensional random vector having multivariate normal distributions with

(2) $$E\epsilon = 0$$

(3) $$E\epsilon\epsilon' = \sigma^2 I_n$$

where $\sigma^2 > 0$ is an unknown parameter and I_n is used to denote the n × n identity matrix.

Biometrics Unit, Cornell University, Ithaca, New York.

The least squares (LS) estimate $\hat{\beta}$ of β and $\hat{\epsilon}$ of ϵ are

$$\hat{\beta} = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2} \quad . \tag{4}$$

and

$$\hat{\epsilon} = Y - X\hat{\beta} \tag{5}$$

$$= Y - X \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$$

$$= PY$$

where

$$P = I_n - \frac{1}{\sum_{i=1}^{n} x_i^2} XX' \quad . \tag{6}$$

It is easy to check that P is a projection (that is, $P = P'$ and $P^2 = P$) and has rank n - 1. Under the above assumptions

$$\begin{aligned} E\hat{\epsilon} &= PX\beta = 0\beta = 0 \\ E\hat{\epsilon}\hat{\epsilon}' &= PEYY'P = PE\epsilon\epsilon'P' = \sigma^2 P \quad . \end{aligned} \tag{7}$$

We see that even **if** assumption (d) is true, the LS estimates of residuals are neither independent nor do they have constant variance since $P \neq I_n$.

Theil [8] has presented an estimator of ϵ which has all the ordinary properties of $\hat{\epsilon}$ except that covariance of Theil estimator is $\sigma^2 I_{n-1}$ under assumption (d). The dimension of Theil estimator of ϵ is n - 1 due to the fact that residual has n - 1 degrees of freedom. While Theil has given the general

procedure for deriving uncorrelated residuals with constant variance under homoscedasticity assumption in multiple linear regression, Koerts [6] has derived the explicit form of the Theil estimator for model (1) which we will use in our diagnostic checking on residual to detect a special type of heteroscedasticity by means of the peak test introduced by Goldfeld and Quandt [3].

Theil Estimator of Residuals

We denote this estimator by ϵ^* and for model (1) it can be represented simply as

$$\epsilon_i^* = y_i - b^* x_i \qquad i = 1, 2, \cdots, k-1, k+1, \cdots, n, \tag{8}$$

where

$$b^* = \left[(1 - a)\hat{\beta}_{n-1} + a \frac{y_k}{x_k} \right] \tag{9}$$

where

$$a = \frac{|x_k|}{\sqrt{\sum_{i=1}^{n} x_i^2}} \quad , \quad \hat{\beta}_{n-1} = \frac{\sum_{\substack{i=1 \\ i \neq k}}^{n} x_i y_i}{\sum_{\substack{i=1 \\ i \neq k}}^{n} x_i^2} \quad . \tag{10}$$

k can take any value from 1 to n and the choice is largely a matter of power with respect to a specific alternative hypothesis.

Properties of Theil Residuals:

1) ϵ_i^* is a linear function of y,

2) $E\epsilon_i^* = 0, \quad i = 1, 2, \cdots, k-1, k+1, \cdots, n,$

3) $\mathrm{Cov}(\epsilon_i^*, \epsilon_j^*) = 0 \quad \text{if } i \neq j$
$= \sigma^2 \quad \text{if } i = j$, $\quad i,j = 1, 2, \cdots, k-1, k+1, \cdots, n,$

4) have the minimum expected sum of squares of estimation residuals in the class of estimators with properties 1, 2, and 3,

5) $$\sum_{\substack{i=1 \\ i\neq k}}^{n} \epsilon_i^{*2} = \sum_{i=1}^{n} \hat{\epsilon}_i^2$$

Properties 3) and 5) make Theil residuals very interesting indeed. Theil residuals have been derived based on the first four properties. Koerts [6] has shown that Theil residuals also have the fifth property. We found it interesting not to use the idea of the derivation of the Theil residuals as the proof, but simply to show that these residuals indeed satisfy the above properties. With this in mind, an appendix at the end of this paper has been devoted to the proof of some of these properties.

Application of Theil Residuals

Consider the case where x's have been ordered such that $x_i < x_j$ if $i < j$. And suppose we are interested in testing the following hypothesis:

$$H_0: \quad \Sigma = \begin{bmatrix} \sigma^2 & & 0 \\ & \ddots & \\ 0 & & \sigma^2 \end{bmatrix}$$

versus

$$H_1: \quad \Sigma = \begin{bmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_n^2 \end{bmatrix}$$

such that $\sigma_i^2 < \sigma_j^2$ for $i < j$,

where Σ stands for the covariance matrix of ϵ. Note that alternative hypothesis says that as x increases variance of ϵ or y increases too. We are considering the case where we have only a single observation for each level of x as it is the case in most experiments.

There are two tests for testing H_0 against H_1.

1) F Test

The obvious choice for k is then the middle observations, so that one can compute the ratio of the sum of squares of the first $\frac{1}{2}(n-1)$ estimated residuals to that of the last $\frac{1}{2}(n-1)$, which is F distributed. When n - 1 is not even, one can use either the $\frac{1}{2}(n-2)$ first and $\frac{1}{2}(n)$ last observations or $\frac{1}{2}(n)$ first and the $\frac{1}{2}(n-2)$ last observations and for this choice see Theil [8].

2) Peak Test

While F test is a general test, peak test has been constructed specially for test of this particular H_0 versus H_1 under our consideration and therefore one expects to obtain a more valid conclusion from this test than F test. Especially when the number of observations is small one prefers to use this test rather than F test because of the small number of degrees of freedom associated with F.

The idea of this test originally has been given by Goldfeld and Quandt [3]. They define a peak for the ordered residuals with respect to x_i, such that $x_i < x_{i+1}$ at observation i, to be an instance where $|\hat{\epsilon}_i| > |\hat{\epsilon}_j|$ for $j = 1, 2, \cdots, i - 1$. Robson and Hedayat [5] have shown the failure of Goldfeld and Quandt for their application of peak test to $\hat{\epsilon}_i$. While application of peak test to $\hat{\epsilon}$ is not strictly valid, its application to ϵ^* is valid and appropriate because

a) under H_0, ϵ_i^*'s are uncorrelated and therefore under the normality assumption will be independent,

b) under H_1, var $\epsilon_i^* (= E\epsilon_i^{*2}) <$ var $\epsilon_{i+1}^* (= E\epsilon_{i+1}^{*2})$ and hence one expects $\epsilon_i^{*2} < \epsilon_{i+1}^{*2}$ or equivalently $|\epsilon_i^*| < |\epsilon_{i+1}^*|$.

Proof of part b) follows from the following theorem.

Theorem:

If

$$\Sigma = \begin{bmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_n^2 \end{bmatrix}$$

such that $\sigma_i^2 < \sigma_j^2$ for $i < j$ and Σ is the covariance matrix of ϵ, then var $\epsilon_i^* <$ var ϵ_j^* for $i < j$.

Proof:

First we compute the variance of ϵ_i^* under the above covariance matrix for ϵ.

$$\text{Var } \epsilon_i^* = E\epsilon_i^{*2} - (E\epsilon_i^*)^2$$

$$= E\epsilon_i^{*2}$$

$$= E\left[y_i - x_i\hat{\beta}_{n-1} - ax_i\left(\hat{\beta}_{n-1} - \frac{y_n}{x_n}\right)\right]^2$$

$$= x_i^2\beta + \sigma_i^2 + \beta^2\left[x_i^2(1 - a)^2 + x_i^2a^2 + 2a(1 - a)x_i^2\right]$$

$$+ \frac{(1-a)^2x_i^2\sum_{i=1}^{n-1} x_i^2\sigma_i^2}{\sum_{i=1}^{n-1} x_i^2} + \frac{x_i^2a\sigma_n^2}{x_n^2}$$

$$- 2x_i(1 - a)\left(x_i\beta^2 + \frac{x_i\sigma_i^2}{\sum_{i=1}^{n-1} x_i^2}\right) - 2x_i^2\beta a$$

$$= \beta^2\left[x_i^2 + x_i^2(1 - a)^2 + x_i^2a^2 + 2a(1 - a)x_i^2 - 2x_i^2(1 - a) - 2x_i^2a\right]$$

$$+ \sigma_i^2 + \frac{(1-a)^2x_i^2\sum_{i=1}^{n-1} x_i^2\sigma_i^2}{\sum_{i=1}^{n-1} x_i^2} + \frac{ax_i^2\sigma_n^2}{x_n^2} - 2x_i(1 - a)\frac{x_i^2\sigma_n^2}{x_n^2}$$

$$= \sigma_i^2 + A^2x_i^2 \sum_{i=1}^{n-1} x_i^2\sigma_i^2 + Bx_i^2 - 2Ax_i^2\sigma_i^2$$

where

$$A = \frac{1-a}{\sum_{i=1}^{n-1} x_i^2} \quad , \quad B = \frac{\sigma_n^2(1-a^2)}{\sum_{i=1}^{n-1} x_i^2}$$

$$= \sigma_i^2(Ax_i^2 + 1)^2 + x_i^2A^2 \sum_{\substack{j=1 \\ j\neq i,i+1}}^{n-1} x_j^2\sigma_j^2 + x_i^2A^2x_{i+1}^2\sigma_{i+1}^2 + Bx_i^2$$

$$= \sigma_i^2(Ax_i^2 + 1)^2 + CA^2x_i^2 + A^2x_i^2x_{i+1}^2\sigma_{i+1}^2 + Bx_i^2$$

where

$$C = \sum_{\substack{j=1 \\ j\neq i,i+1}}^{n-1} x_j^2\sigma_j^2 \ .$$

Now

$$\text{Var } \epsilon_{i+1}^* - \text{Var } \epsilon_i^* = \sigma_{i+1}^2(Ax_{i+1}^2 + 1)^2 - \sigma_i^2(Ax_i^2 + 1)^2 + CA^2(x_{i+1}^2 - x_i^2)$$

$$+ A^2x_i^2x_{i+1}^2(\sigma_i^2 - \sigma_{i+1}^2) + B(x_{i+1}^2 - x_i^2) \ .$$

For $x_{i+1} > x_i$ we have

$$CA^2(x_{i+1}^2 - x_i^2) > 0 \quad \text{and} \quad B(x_{i+1}^2 - x_i^2) > 0 \ .$$

Now in order to show that var ϵ_{i+1}^* > var ϵ_i^* we have to show that

$$\sigma_{i+1}^2(Ax_{i+1}^2 + 1)^2 - \sigma_i^2(Ax_i^2 + 1) + A^2x_i^2x_{i+1}^2(\sigma_i^2 - \sigma_{i+1}^2) > 0 \ .$$

That is,

$$\sigma_{i+1}^2A^2x_{i+1}^2x_{i+1}^2 - \sigma_i^2A^2x_i^2x_i^2 + \sigma_{i+1}^2(1 + 2Ax_{i+1}^2)$$

$$- \sigma_i^2(1 + 2Ax_i^2) + A^2x_i^2x_{i+1}^2(\sigma_i^2 - \sigma_{i+1}^2) > 0 \ .$$

Since $x_i < x_{i+1}$, then the left side of above final expression will be greater than

$$\sigma_{i+1}^2 A^2 x_i^2 x_{i+1}^2 - \sigma_i^2 A^2 x_i^2 x_{i+1}^2 + \sigma_{i+1}^2(1 + 2Ax_{i+1}^2)$$

$$- \sigma_i^2(1 + 2Ax_i^2) + A^2 x_i^2 x_{i+1}^2(\sigma_i^2 - \sigma_{i+1}^2)$$

$$= \sigma_{i+1}^2(1 + 2Ax_{i+1}^2) - \sigma_i^2(1 + 2Ax_i^2) > 0 .$$

Q.E.D.

A Numerical Example

We apply peak test using the Theil residuals to the example given by Steel and Torrie [7] on page 180 of their textbook for the purpose of regression through the origin.

Induced Reversions to Independence per 10^7 Surviving Cells y per Dose (ergs/Bacterium) 10^{-5}x of Stroptomycin Dependent Escherichia Coli Subjected to Monochromatic Ultraviolet Radiation of 2,967 Angstroms Wavelength.

x	y
13.6	52
13.9	48
21.1	72
25.6	89
26.4	80
39.8	130
40.1	139
43.9	173
51.9	208
53.2	225
65.2	259
56.4	199
67.7	255
$\sum_{i=1}^{13} x_i = 528.8$	$\sum_{i=1}^{13} y_i = 1,929$
$\sum_{i=1}^{13} x_i^2 = 26,062.10$	$\sum_{i=1}^{13} y_i^2 = 356,259$

Regression of y on x should pass through the origin as Steel and Torrie [7] have shown. Therefore,

$$\hat{\beta} = \frac{\sum_{i=1}^{13} x_i y_i}{\sum_{i=1}^{13} x_i^2} = 3.67$$

and hence the regression line is given by

$$y = 3.67x \ .$$

The reduction in sum of squares attributable to regression is

$$\left(\sum_{i=1}^{13} x_i y_i \right)^2 / \sum_{i=1}^{13} x_i^2 = 351,819.$$

And the residual sum of squares is 356,259 - 351,819 = 4,440.

The individual least square residuals are:

$$\begin{aligned}
\hat{\epsilon}_1 &= + 2.088 \\
\hat{\epsilon}_2 &= - 3.013 \\
\hat{\epsilon}_3 &= - 5.437 \\
\hat{\epsilon}_4 &= - 4.952 \\
\hat{\epsilon}_5 &= - 16.888 \\
\hat{\epsilon}_6 &= - 16.066 \\
\hat{\epsilon}_7 &= - 8.167 \\
\hat{\epsilon}_8 &= + 11.887 \\
\hat{\epsilon}_9 &= + 17.527 \\
\hat{\epsilon}_{10} &= + 29.756 \\
\hat{\epsilon}_{11} &= + 19.716 \\
\hat{\epsilon}_{12} &= - 44.688 \\
\hat{\epsilon}_{13} &= + 6.541 \ .
\end{aligned}$$

First of all, it seems there is a pattern for the distribution of plus and minus signs of $\hat{\epsilon}_i$'s. Second, one gets the impression that it seems the absolute value of $\hat{\epsilon}_i$'s increases as i increases. Now suppose we suspect about the assumptions which we made on page 1. And suppose that the only alternative hypothesis of interest to us is that, variance of ϵ (or y) increases as x increases. To test this hypothesis first we find the Theil residuals

$$\epsilon_i^* = y_i - b^* x_i \quad , \quad i = 1, 2, \cdots, 12$$

where

$$b^* = \left[(1 - a)\hat{\beta}_{12} + a \frac{y_{13}}{x_{13}} \right]$$

$$= \left(1 - \frac{|x_{13}|}{\sqrt{\sum_{i=1}^{13} x_i^2}} \right) \frac{\sum_{i=1}^{12} x_i y_i}{\sum_{i=1}^{12} x_i^2} + \frac{|x_{13}|}{\sqrt{\sum_{i=1}^{13} x_i^2}} \frac{y_{13}}{x_{13}}$$

$$= \left(1 - \frac{67.7}{\sqrt{26,062.10}} \right) \frac{78,492.2}{21,478.81} + \frac{67.7}{\sqrt{26,062.10}} \frac{255}{67.7}$$

$$= 3.7014.$$

Therefore,

$$\epsilon_i^* = y_i - 3.7014 x_i \quad , \quad i = 1, 2, \cdots, 12 .$$

Hence,

$$\epsilon_1^* = + 1.66096$$

$$\epsilon_2^* = - 3.44946$$

$$\epsilon_3^* = - 6.09954$$

$$\epsilon_4^* = - 5.75584$$

$$\epsilon_5^* = - 17.71696$$

$$\epsilon_6^* = - 17.31572$$

$$\epsilon_7^* = - 9.42614$$
$$\epsilon_8^* = + 10.50854$$
$$\epsilon_9^* = + 15.89734$$
$$\epsilon_{10}^* = + 28.08552$$
$$\epsilon_{11}^* = + 17.66872$$
$$\epsilon_{12}^* = - 46.77296 \qquad \text{No. of peaks} = 5$$

$\sum_{i=1}^{12} \epsilon_i^{*2} = 4,439.4133$ and slight departure from 4,440 is due to the rounding error that we have done in computing b^*.

ϵ_i^*'s are independent and identically distributed under homoscedasticity and normality assumption of ϵ_i's. Now we can compute the probability of obtaining five peaks in a sequence of 12 independent and identically distributed random variables. We can use the following table which has been computed by Goldfeld and Quandt [3] for this purpose.

CUMULATIVE PROBABILITIES FOR THE DISTRIBUTION OF PEAKS

	P (number of peaks ≤ x)										
n	x = 0	x = 1	x = 2	x = 3	x = 4	x = 5	x = 6	x = 7	x = 8	x = 9	x = 10
5	.2000	.6167	.9083	.9917	1.0000						
10	.1000	.3829	.7061	.9055	.9797	.9971	.9997	1.0000			
15	.0667	.2834	.5833	.8211	.9433	.9866	.9976	.9997	1.0000		
20	.0500	.2274	.5022	.7530	.9056	.9720	.9935	.9988	.9998	1.0000	
25	.0400	.1910	.4441	.6979	.8705	.9559	.9879	.9973	.9995	.9999	1.0000
30	.0333	.1654	.4001	.6525	.8386	.9395	.9815	.9953	.9990	.9998	1.0000
35	.0286	.1462	.3654	.6144	.8098	.9234	.9745	.9929	.9984	.9997	.9999
40	.0250	.1313	.3373	.5818	.7837	.9078	.9674	.9903	.9975	.9995	.9999
45	.0222	.1194	.3138	.5536	.7600	.8930	.9601	.9874	.9966	.9992	.9998
50	.0200	.1096	.2940	.5288	.7383	.8788	.9530	.9844	.9956	.9989	.9998
55	.0182	.1014	.2769	.5068	.7184	.8653	.9456	.9813	.9944	.9986	.9997
60	.0167	.0944	.2620	.4871	.7001	.8524	.9384	.9780	.9932	.9982	.9996

By interpolation from this table we see that the probability is about .06 that a sequence of 12 independent and identically distributed random variables produce five peaks. If we can accept a risk of 6 percent and if our suspicion about homoscedasticity has biological support, then we should fit the weighted regression rather than the unweighted one for obtaining an efficient estimate of β and hence the regression line.

Conclusion

If we pay no heed to checking the textbook ideal assumptions related to the linear model, this means either we are willing to accept these assumptions, or simply that we are not aware of the importance of these basic assumptions. But truly there is no reason to suppose that the conventional assumptions are ever satisfied in practice. Therefore, we would like to have method(s) for detecting and measuring any sort of departure from these ideal conditions.

Examination and analysis of residuals is concerned with detecting and measuring certain sorts of departures from the conventional assumptions about the linear model. As Anscombe and Tukey [2] have put it, "If we are to improve our analysis of data to which the conventional techniques can be applied, it is not likely that we shall do this by improving the techniques themselves. Rather we must learn either to go further, beyond the place where the conventional techniques stop, or we must learn to use the techniques better. Either path demands the analysis of residuals, where

$$(\text{residual}) = (\text{observed value}) - (\text{fitted value}).$$

In the first path we analyze residuals to learn what they can tell us of direct interest. In the second path we must analyze the residuals from a first application of conventional methods to learn how a second application might be better made."

Least squares residuals, even if under the ideal conditions, are in general correlated and have different variances. We think perhaps for graphical examination of residuals in certain cases we can neglect both covariance and heterogeneity of variances which exist among the least squares residuals. But, certainly for constructing tests or any rigorous examination of residuals, we prefer to work with a new type of residuals which are free from the above criticism.

Surely, as we learn to do better data analysis, computation will get more extensive rather than simpler. But, if sophisticated data analysts are to gain in depth and power, they must have both the time and stimulation to try out new procedures of analysis. I take advantage here and I quote Tukey's [9] words of wisdom, "The future of data analysis can involve great progress, the overcoming of real difficulties, and the provision of a great service to all fields of science and technology. Will it? That remains to us, to our willingness to take up the rocky road of real problems in preference to the smooth road of unreal assumptions, arbitrary criteria, and abstract results without real attachments. Who is for the challenge?"

Appendix

Proof of properties of Theil residuals:

1) ϵ_i^* is a linear function of y's.

   Proof follows by definition of ϵ_i^*.

2) $E\epsilon_i^* = 0$. This is so because

$$E\epsilon_i^* = Ey_i - x_i Eb^*$$

$$= x_i\beta - x_i \left[(1 - a)\beta + a \frac{x_k\beta}{x_k}\right] = 0$$

3) $\mathrm{Cov}(\epsilon_i^*, \epsilon_j^*) = 0 \quad \text{if } i \neq j$

$$= \sigma^2 \quad \text{if } i = j$$

To prove this we write n - 1 Theil residuals in vector notation as

$$\epsilon^* = \left[\frac{-a}{x_k} X_1' \quad I - \frac{1-a}{X_1'X_1} X_1X_1' \right] Y ,$$

where X_1 is the column vector X from which we have taken component x_k out. Then the covariance matrix of ϵ^* will be

$$E(\epsilon^*\epsilon^{*\prime}) = E \left[\frac{-a}{x_k} X_1' \quad I - \frac{1-a}{X_1'X_1} X_1X_1' \right] \epsilon\epsilon' \begin{bmatrix} \frac{-a}{x_k} X_1 \\ I - \frac{1-a}{X_1'X_1} X_1X_1' \end{bmatrix}$$

$$= \sigma^2 \left[\frac{-a}{x_k} X_1' \quad I - \frac{1-a}{X_1'X_1} X_1X_1' \right] \begin{bmatrix} \frac{-a}{x_k} X_1 \\ I - \frac{1-a}{X_1'X_1} X_1X_1' \end{bmatrix}$$

$$= \sigma^2 \left[\frac{a^2}{x_k^2} X_1X_1' + I - 2 \frac{1-a}{X_1'X_1} X_1X_1' + \frac{(1-a)^2}{X_1'X_1} X_1X_1' \right]$$

$$\text{since} \quad a^2 = \frac{x_k^2}{X'X} = \frac{x_k^2}{X_1'X_1 + x_k^2}$$

$$\Rightarrow x_k^2 = \frac{a^2 X_1'X_1}{1-a^2}$$

$$= \sigma^2 \left[\frac{1-a^2}{X_1'X_1} X_1X_1' + I - 2 \frac{1-a}{X_1'X_1} X_1X_1' + \frac{(1-a)^2}{X_1'X_1} X_1X_1' \right]$$

$$= \sigma^2 \left[I \right] .$$

4) Have the minimum expected sum of squares of estimation residuals in the class of estimators with properties 1, 2, and 3. For the proof see Theil [8].

5) $$\sum_{\substack{i=1 \\ i\neq k}}^{n} \epsilon_i^{*2} = \sum_{i=1}^{n} \hat{\epsilon}_i^2 \quad .$$

The proof is as follows:

$$\sum_{\substack{i=1 \\ i\neq k}}^{n} \epsilon_i^{*2} = \epsilon^{*\prime}\epsilon^{*} \quad \text{in vector notation}$$

$$= Y' \begin{bmatrix} \frac{-a}{x_k} X_1' \\ I - \frac{1-a}{X_1'X_1} X_1 X_1' \end{bmatrix} \begin{bmatrix} \frac{-a}{x_k} X_1 & I - \frac{1-a}{X_1'X_1} X_1 X_1' \end{bmatrix} Y$$

$$= Y' \begin{bmatrix} \frac{a^2 X_1'X_1}{x_k^2} & \frac{-a^2}{x_k} X_1' \\ \frac{-a^2}{x_k} X_1 & I + \frac{a^2-1}{X_1'X_1} X_1 X_1' \end{bmatrix} Y$$

$$\text{since } a^2 = \frac{x_k^2}{X'X} \quad \text{and} \quad x_k^2 = X'X - X_1'X_1$$

$$= Y' \begin{bmatrix} 1 - \frac{x_k^2}{X'X} & \frac{-x_k}{X'X} X_1 \\ \frac{-x_k}{X'X} X_1 & I - \frac{1}{X'X} X_1 X_1' \end{bmatrix} Y$$

$$= Y' \left[I - \frac{1}{X'X} XX' \right] Y$$

$$= Y'PY$$

$$= Y'P'PY$$

$$= \hat{\epsilon}'\hat{\epsilon} \ .$$

References

[1] Anscombe, F. J. "Examination of Residuals." Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics, I:1-36, 1960.

[2] Anscombe, F. J. and J. W. Tukey. "The Examination and Analysis of Residuals." Technometrics 5:141-159, 1963.

[3] Goldfeld, S. M. and R. E. Quandt. "Some Test for Homoscedasticity." Journal of the American Statistical Association 60:539-547, 1965.

[4] Hedayat, A. "Homoscedasticity in Linear Regression Analysis with Equally Spaced X's." M. S. Thesis, Cornell University, Ithaca, New York, June, 1966.

[5] Hedayat, A. and D. S. Robson. "Independent Transformed Residuals for Testing Homoscedasticity." Paper No. BU-135, Biometrics Unit, Cornell University, Ithaca, New York, 1966.

[6] Koerts, J. "Some Further Notes on Disturbance Estimates in Regression Analysis." Journal of the American Statistical Association 62:169-183, 1967.

[7] Steel, R. G. D. and J. H. Torrie. "Principle and Procedures of Statistics." McGraw-Hill Book Company, Inc., New York, 1960.

[8] Theil, H. "The Analysis of Disturbances in Regression Analysis." Journal of the American Statistical Association 60:1067-1079, 1965.

[9] Tukey, J. W. "The Future of Data Analysis." The Annals of Mathematical Statistics 33:1-67, 1962.